

R E P O R T R E S U M E S

ED 019 197

RE 001 219

THE ASSESSMENT OF CHILDREN'S STATEMENTS OF THE MAIN IDEA IN
READING.

BY- OTTO, WAYNE AND OTHERS

PUB DATE 26 APR 68

EDRS PRICE MF-\$0.25 HC-\$0.68 15P.

DESCRIPTORS- *READING COMPREHENSION, *RATING SCALES, *READING
TESTS, READABILITY, VERBAL STIMULI, *FUNCTIONAL READING,

A CONTINUATION OF A STUDY OF PUPILS' ABILITY TO FORMULATE AND STATE A LITERAL MAIN IDEA IN THE READING OF SHORT, SPECIALLY CONSTRUCTED PARAGRAPHS IS REPORTED. ONE OF THE MAJOR TASKS WAS TO DEVELOP A DESCRIPTIVE PROFILE OF THE MAIN IDEA RESPONSES FOR A GROUP OF 400 SECOND- AND FIFTH-GRADE PUPILS. A 12-POINT SCALE WAS USED FOR THE NUMERICAL ORDERING OF RESPONSES OF THREE PARAGRAPHS AT THE FIRST-GRADE READABILITY LEVEL. RESPONSES REQUIRED A SYNTHESIS OF BOTH SUBJECT AND PREDICATE IN ONE SENTENCE AND WERE CODED AND CATEGORIZED BY THREE JUDGES. DATA FROM BOTH GRADES WERE EXAMINED SEPARATELY TO MAKE INFORMAL COMPARISONS. INTERJUDGE RELIABILITY COEFFICIENTS WERE VERY HIGH FOR PARAGRAPHS COMBINED AND SEPARATE. LOW INTERPARAGRAPH CORRELATIONS INDICATED THE PARAGRAPHS WERE NOT INTERCHANGEABLE. SECOND GRADERS TENDED TO REPLY IN TITLE-LIKE RESPONSES, INDICATING AN INABILITY TO COPE WITH COMPLETE STATEMENTS. SINCE THE MAJORITY OF FIFTH GRADERS WERE ABLE TO RESPOND ADEQUATELY, THERE IS A NEED FOR WORK WITH MORE COMPLEX TASKS AT THIS LEVEL. FURTHER EXPERIMENTATION IS PLANNED. REFERENCES ARE LISTED. THIS PAPER WAS PRESENTED AT THE INTERNATIONAL READING ASSOCIATION CONFERENCE (BOSTON, APRIL 24-27, 1968). (MC)

Wayne Otto
The University of Wisconsin
1815 University Avenue
Madison, Wisconsin 53706

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE
PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION
POSITION OR POLICY.

INTERNATIONAL READING ASSOCIATION, BOSTON, APR. 24-27, 1968,

The Assessment of Children's Statements
of the Main Idea in Reading¹

Wayne Otto, Thomas C. Barrett and Karl Koenke
The University of Wisconsin, Madison
(Research Reports: Comprehension, Friday, April 26, 1:30-2:30 p.m.)

The results of preliminary work designed to establish a framework for the study of elementary pupils' ability to formulate and state a literal main idea in reading has been reported elsewhere (Barrett & Otto, 1968; Otto & Barrett, in press). Because specific methodological and conceptual precedents did not exist, the focus in the earlier work was upon (a) methodological matters, mainly devising and testing materials, procedures and response scaling methods appropriate for use with

¹The research reported herein was partially supported by a contract with the United States Office of Education, Department of Health, Education and Welfare, under the provisions of the Cooperative Research Program. Center No. C-03/Contract OE 5-10-154.

elementary school children, and (b) descriptive data derived from analyses of children's attempts to formulate and state the main ideas implicit but not explicitly stated in short, specially constructed paragraphs. In the present study the two main purposes were to (a) refine the scale and the judging procedures used to assign quantitative ratings to individual's main idea statements, and (b) make use of the newly developed material and procedures in seeking additional descriptive data regarding children's main idea statements.

One of the major tasks that we confronted in our earlier attempts to examine the nature of children's main idea statements was to devise a means for scoring or categorizing the statements for purposes of description and analysis. The final decision was to develop a scale that could be used to place responses into descriptive categories and to quantify response quality. After much exploratory work and revision we developed a six-point scale, with category values ranging from 0, where there was no response, to 6, where the main idea statement was adequate in terms of both general and restrictive elements. The six-point scale was used in our exploratory study, but it became clear in application that provision of finer distinctions would permit more accurate judgments and more explicit descriptions of responses. Consequently, a twelve-point scale, referred to hereafter as SNORT (Scale for Numerical Ordering of Response Tendencies), was devised and used in the present study.

With the decision to employ a new scale in evaluating main idea responses, it became necessary to re-examine the reliability of ratings by judges who used the scale. With the provision of more explicit

categories, the expectation was that inter-judge reliability would increase; but the possibility that finer distinctions would lead to increased confusion remained. One specific focus of the present study, then, was upon validation of SNORT in terms of inter-judge reliability.

A second specific focus was upon the question of interchangeability of stimulus paragraphs. In our earlier work we devised three paragraphs that were comparable in readability ratings, length, linguistic constraints and information content. We assumed that the paragraphs were equal--that is, that children would be equally successful in extracting, formulating and stating main ideas for each of them--but we worked only with the sums of scores from all three paragraphs. Thus, the question of whether the paragraphs might not in fact differ was neither answered nor asked. In the present study we did ask the question because if the paragraphs are indeed equal there would be little reason to work with more than one in future studies.

A final focus of the study was upon the SNORT ratings of the main idea responses of 400 elementary school children. The intent was to create a descriptive profile of the main idea responses of a fairly large sample of children and to examine it for implications.

Method

Subjects and Procedure

Subjects were chosen from among the 512 second and fifth grade pupils enrolled in four Madison, Wisconsin, elementary schools. Children were chosen at random from the two grade levels and asked to read a sample test

selection. Only those who had no difficulty with the mechanics of the reading task were retained as subjects. Two hundred subjects, equal numbers of boys and girls, from each grade level were tested, for a total of 400 subjects.

Each subject, tested individually in a private room in the school building, read and attempted to state the main idea of three paragraphs. The directions and procedure are described elsewhere (Otto & Barrett, in press) in detail. Briefly, each testing session began with a short warm-up task in which the subject was asked to compose and read back four simple sentences. The purpose of the warm-up was to establish both rapport and a set to respond in complete sentences. Then the subject was asked to read a paragraph silently while thinking about "what all the sentences say together." Pilot testing showed that the latter phrase was effective in directing the subject toward a synthesis of all the elements rather than selection of a single specific thought. The examiner supplied any words questioned by the subject. When the reading was completed, the subject was told to "make up just one sentence in your own words that says what all the sentences tell you." The directions were partially repeated between paragraphs. Order of presentation was systematically varied so that each paragraph was read first, second and third an approximately equal number of times. The entire task took approximately eight minutes.

Material

Details of the underlying rationale and the development of the paragraphs used in our earlier studies are given elsewhere (Otto & Barrett, in press). Paragraphs for those studies were developed at six readability levels, ranging from grade one to grade six, and in two styles, one requiring synthesis of both subject and predicate and the other requiring synthesis of the predicate only for an adequate main idea statement. The three paragraphs chosen for use in the present study (a) had a first grade readability rating because we wanted to minimize the mechanical demands of the reading task, and (b) required synthesis of both subject and predicate, although our previous work revealed no difference between the two styles studied. The actual paragraphs follow:

Paragraph A

Cats help the farmer keep mice from his corn. A horse helps the farmer work. Cows give milk to the farmer. A dog helps the farmer watch the barnyard. (Main idea: Animals help farmers in different ways.)

Paragraph B

Robins may build nests under a roof. Bluejays like nests in trees. Ducks make nests in tall grass. Woodpeckers make nests inside wood fence posts. (Main idea: Birds build nests in different places.)

Paragraph C

Lions use claws to hold their food. Bears have claws for digging. Cats' claws help them climb trees quickly. Tigers use strong claws for killing. (Main idea: Animals use their claws for different purposes.)

Response Scale

The twelve-point scale devised for use in the present study is presented in the schema that follows. An operational definition is given for each category.

Scale for Numerical Ordering of
Response Tendencies (SNORT)
With Examples

<u>Category and Examples</u>	<u>Scale Value</u>
No response, or a response showing that the task was not understood, i.e., reading stimulus material. 1. "About...." Repeated question twice. "I don't know." 2. Read orally a sentence from the paragraph. 3. Read the paragraph orally.	0
Irrelevant or incorrect material, i.e., paraphrasing or bring in material not directly related to the paragraph. 1. Rabbits can be pets. 2. Robins make nests in trees. 3. Baby chicks need food.	1
One element partially given. 1. This story is talking about birds. 2. Nests. 3. Many nests are made.	2
One element too generally or specifically stated, plus irrelevant or incorrect data. 1. Birds, nests, and fences. 2. It says robins, bluejays, woodpeckers, and ducks make nests in the grass. 3. Some bluejays, ducks, and robins make <u>nests</u> and stuff... that's all.	3
One element too generally or specifically stated. 1. Robins, ducks, bluejays have nests. 2. Birds all build something. 3. Birds in their nests.	4

- Two elements too generally or specifically stated plus irrelevant or incorrect data. 5
1. It tells about birds, the kind, what they do, where they build their bird houses, what they build it out of. (Only one example given)
- Two elements too generally or specifically stated. 6
1. Where bluejays, birds, and ducks put their nests.
 2. Robins, bluejays, and woodpeckers make their nests in high grass, under rooftops and in fence posts.
 3. What birds make their places in.
- One element correctly stated plus irrelevant or incorrect data. 7
1. How birds make nests.
 2. Birds make nests in trees.
 3. Different kinds of birds make different kinds of nests.
- One element correctly stated. 8
1. Birds make nests.
 2. Birds are nesting.
 3. The birds all make a home.
- One element correctly stated plus one element too generally or specifically stated plus irrelevant or incorrect data. 9
1. How and where birds make their nests.
 2. What kind of birds make what kind of nest and where.
 3. That birds build nests in some places that you wouldn't build your house in.
- One element correctly stated plus one element too generally or specifically stated. 10
1. Where birds like to build nests.
 2. Robins, bluejays, and ducks build nests in different places.
 3. The birds make nests in things or under them.
- Two elements correctly stated plus irrelevant or incorrect data. 11
1. Birds make different nests in different places.
 2. Some birds build nests in different places.
 3. That all birds make nests and some make nests in different places.
- Two elements correctly stated. 12
1. Birds make nests in different places.
 2. Birds build nests in many places.
 3. Different birds build nests in different places.

In effect, main idea responses are ranked on SNORT by the degree of synthesis within each response. Optimal value is placed upon appropriate synthesis of both subject and predicate elements, over- or undergeneralizations of either element have a lesser value, and parroted specifics or irrelevant associations have little or no value. Application of the scale is limited to literal main idea statements inferred from the explicit content of the paragraphs.

Scoring Responses

Each subject's main idea responses were coded, scrambled and typed on master sheets prior to judging. Three judges, who had participated in training sessions devoted to rating sample responses on SNORT, independently assigned each response to a SNORT category. Thus, there were three ratings for each of the three responses given by individual subjects. The judges met three times during the time that they were assigning ratings to discuss their individual operational approaches to the task in order to keep idiosyncratic percepts and interpretations from running wild.

Results and Discussion

Data from the second and fifth grade subjects were examined separately. As expected, we found in our earlier work that children's main idea responses increase in quality across grade level. In the present study we simply wanted to make informal comparisons of data from primary and intermediate grade children.

Interjudge Agreement

One of our concerns was whether there would be high agreement among judges using SNORT to rate main idea statements. When the ratings for all three paragraphs were totaled, the interjudge reliability coefficients were .975, .973 and .974 with second grade data, and .962, .962 and .965 with fifth grade data. Likewise, interjudge reliability coefficients for each paragraph taken individually were in all cases higher than .91. Agreement among judges, then, was very high when ratings for all paragraphs combined and for single paragraphs were considered. On the basis of this finding, it would appear to be defensible to place confidence in the ratings of a single judge working with SNORT.

Comparison of Paragraphs

A second concern was whether our three carefully constructed paragraphs were, as we hoped, interchangeable. To answer this question we first examined intercorrelations for main idea ratings for each paragraph. If the three paragraphs evoked main idea statements of comparable SNORT ratings, then the intercorrelations among them would be expected to be about as high as the intercorrelation among judges' ratings of each paragraph. With the second grade data, the intercorrelations of the sum of the three judges' ratings for each paragraph were: A-B = .562, A-C = .591, B-C = .517. With fifth grade data the intercorrelations were: A-B = .248, A-C = .306, B-C = .531. At each grade level, the intercorrelations of individual judge's ratings of the paragraphs were almost identical in magnitude to those given, which again

supports the suggestion that a single judge serves as well as three. The low interparagraph correlations show that the three paragraphs, though related, do not appear to be interchangeable despite the controls imposed when they were written. ✓

The possibility remained that the order in which the paragraphs were presented may have had some effect upon the responses evoked, due either to a straightforward training effect or to an interaction between content and order. Although the order of presentation was varied to break up any systematic order effect, the number of subjects who read each paragraph in each order was not exactly equal because some improvising had been done to round out numbers of subjects in the four schools. We decided to run an analysis of variance to examine Order and Paragraph effects and the Order x Paragraph interaction; however, we first set the number of subjects who read each paragraph in each order equal and were left with 183 second grade and 192 fifth grade subjects. Mean SNORT ratings of these subjects' responses are given by order and paragraph in Table 1.

Table 1

Mean SNORT Ratings
by Paragraph and Order

	Grade							
	2 (N = 183)				5 (N = 192)			
	Order				Order			
Paragraph	1	2	3	Σ	1	2	3	Σ
A	4.23	4.11	4.38	4.24	7.22	8.22	7.70	7.80
B	5.61	5.51	5.84	5.65	8.70	8.34	9.64	8.90
C	4.11	3.56	4.10	3.92	6.86	8.69	8.03	7.86
Σ	4.65	4.39	4.77	4.60	7.59	8.42	8.55	8.19

Row sums are mean ratings by paragraph and column sums are mean ratings across paragraphs in each order position. The analysis of variance, in which row and column scores were arbitrarily treated as independent, is summarized in Table 2.

Table 2
Analysis of Variance of SNORT Ratings
by Paragraph and Order

		Grade					
		2			5		
		df	MS	F	df	MS	F
Order	(0)	2	6.79	.484	2	51.28	5.21*
Paragraph (P)		2	154.63	11.030*	2	72.76	7.39*
O x P		4	1.06	.076	4	24.82	2.52**
Error		540	14.09	---	575	9.84	---

* $p < .01$

** $p < .05$

With second grade data, only the paragraph effect was significant ($p < .01$). Inspection of Table 1 shows that a relatively higher mean was attained with Paragraph B than with Paragraph A or C. With the fifth grade data, the Order and the Paragraph effects ($p < .01$) as well as the interaction ($p < .05$) were significant. Inspection of the means in Table 1 shows that in general (a) Paragraph B again evoked the most highly rated

main idea statements, and (b) the paragraph presented first evoked the least highly rated main idea statements. As for the interaction, Sheffé tests revealed the Paragraph B statements differed significantly from the Paragraph A and C statements in both first and third position, but there were no differences in second position.

Three conclusions seem justified. First, subjects from both grade levels were able to give the most highly rated main idea statements in response to Paragraph B, which appears to indicate that despite the linguistic constraints the content of that paragraph could most readily be handled in terms of the main idea task. Perhaps the subjects' success with Paragraph B is attributable to differences in their contact with the concepts underlying the three main ideas. That is, second and fifth grade students in metropolitan area schools are more likely to have observed birds building nests than animals working on a farm or using their claws in the wild. The message may be that concrete experiences with concepts explicitly enhance a child's ability to infer a complex main idea. Second, in view of the significant order effect for fifth grade only, it appears that the relatively more sophisticated subjects benefited from practice with the task. The second graders, whose statements were consistently rated relatively low apparently gained nothing from experience. Perhaps the implication is that instruction designed to help pupils derive and state a literal main idea in reading will be more efficient if it is deferred until later in the elementary school experience. This idea, of course, needs to be examined much more

explicitly before final conclusions are reached. Third, if a single paragraph were to be chosen for future work, Paragraph B might be the best choice not only because the subjects were able to respond most adequately to it, but also because the concepts involved may be more familiar to more second and fifth graders.

SNORT Ratings

A third and final focus of the study was upon the actual SNORT ratings of the second and fifth grade subjects' main idea statements. A frequency distribution of the SNORT ratings of subjects' main idea responses to Paragraph B is given in Table 3. Examples of actual responses given to Paragraph B are included with the SNORT scale description in the method section of this paper.

In our earlier work a developmental trend toward higher response ratings across grade levels was clear, and the trend is again clear when the distributions of second and fifth graders' responses are compared. Furthermore, if SNORT categories are collapsed it is possible to say that 28 per cent of the second grade subjects (Categories 0 and 1) could not cope with the task; whereas, this was true of only 2 per cent of the fifth grade subjects. Forty-three per cent of the second grade subjects were bound to the first or a single element of the main idea (Categories 3 to 8). In general these responses tended to be phrases or sentence fragments, similar to the "titles" called for in many traditional comprehension tests. Only 34 per cent of the fifth graders responded in such a fashion. At the upper end of the scale, 29 per cent of the second grade subjects received ratings (9-12) that denoted a response at least approximating the whole main idea sentence;

Table 3

Frequency Distribution of Subjects'
Paragraph B Responses by SNORT Categories

SNORT Category	Grade			
	2		5	
	N	%	N	%
0	9	4.5	0	0
1	47	23.5	3	1.5
2	15	7.5	11	5.5
3	6	3	4	2
4	12	6	5	2.5
5	1	0.5	0	0
6	3	1.5	5	2.5
7	19	9.5	29	14.5
8	30	15	15	7.5
9	5	2.5	11	5.5
10	43	21.5	61	30.5
11	2	1	14	7
12	8	4	42	21

whereas, 64 per cent of the fifth graders' responses received similar ratings. Perhaps, then, second graders have developed a set to respond to a main idea task with title-like statements due to their past experience; or, the task may demand a level of maturity not yet attained by most second graders. Work

with alternate instructional approaches will help to clarify the source of their difficulty. While the majority of fifth graders appear to be able to cope with the task of formulating the main idea when materials are extremely simple and straightforward, work with more complex main idea tasks is needed.

We feel that we have made some progress in tackling a problem that is as complex as it is important. We intend to do more.

References

- Barrett, T. C., and Otto, W. "Elementary Pupils' Ability to Conceptualize the Main Idea in Reading." Paper read at the American Educational Research Association annual meeting, Chicago, 1968.
- Otto, W., and Barrett, T. C. Two studies of children's ability to formulate and state a literal main idea in reading. Technical Report of the Wisconsin Research and Development Center for Cognitive Learning, in press.